

# Large-Scale Data-Driven Financial Risk Modeling using Big Data Technology

Kurt Stockinger, Jonas Heitz, Nils Bundi, Wolfgang Breymann  
Zurich University of Applied Sciences  
Winterthur, Switzerland

**Abstract**—Real-time financial risk analytics is very challenging due to heterogeneous data sets within and across banks worldwide and highly volatile financial markets. Moreover, large financial organizations have hundreds of millions of financial contracts on their balance sheets. Since there is no standard for modelling financial data, current financial risk algorithms are typically inconsistent and non-scalable. In this paper, we present a novel implementation of a real-world use case for performing large-scale financial risk analytics leveraging Big Data technology. Our performance evaluation demonstrates almost linear scalability.

**Index Terms**—Big Data, Parallel Processing, Computational Finance, Data Modeling

## I. INTRODUCTION

In 2008, when Lehman Brothers went bankrupt at the height of the financial crisis, nobody — neither the big banks nor the regulatory authorities — were able to measure, even imprecisely, the risks inherent in the development of subprime lending, securitization, and risk transfer [6]. The Basel Committee on Banking Supervision (BCBS) [2] identified the major shortcoming to be the banks' poor capability to "*quickly and accurately*" aggregate risk exposures and identify concentrations of risk at "*the bank group level, across business lines and between legal entities*". In particular, this affects their ability to quickly carry out so-called stress tests, which can be viewed as a particular way to assess the banks' exposure to risk associated with particular economic-financial scenarios.

In order to achieve the required efficiency, reliability and transparency for performing large-scale risk analysis, one needs (i) an adequate standardization of the data defining the contract and the cash flow generating algorithms [5] and (ii) sufficient computational power. The first goal has already been achieved with the Algorithmic Contract Type Unified Standard (ACTUS) [1], which is a standardization of financial contract data and algorithms based on ideas originally presented in [3], while (ii) is the subject of this paper.

ACTUS builds on the observation that all financial analyses can be reduced to the following three steps: (1) evaluation of the cash flows of all financial contracts for a given state of the economic-financial environment (henceforth referred to as risk-factor scenarios); (2) application of specific analytical transformations to these cash flows, which results in granular analytics; and (3) aggregation of the granular analytics to any desired level. If data and contract algorithms are standardized, steps (2) and (3) consist nearly entirely of linear transformations, for which very efficient parallel algorithms exist.

In order to scale up the solutions, it is important to notice that steps (1) and (2) can be processed independently for different contracts and different risk factor scenarios, which means that the computing problem is "embarrassingly parallel". This property allows leveraging novel and freely available parallel and/or distributed computing technologies in order to solve the ACTUS cash flow generating and analytical algorithms and, consequently, the problem of real-time financial risk analytics. In particular, this is essential for large financial institutions with about  $10^8$  contracts and for Monte-Carlo simulation based analytics where each instrument has to be evaluated under up to  $10^5$  scenarios, thus resulting in a Big Data and Big Computation problem.

## II. ACTUS FRAMEWORK

Figure 1 shows the data flow of the ACTUS framework [4]. The data input required for the simulation step essentially consists of **contract data** and **risk factor scenarios**:

- *Contract data* provide the contractual terms defined in the legal agreements; they determine types, dates, and amount of cash flows generated by a contract.
- *Risk factors scenarios* determine the state of the financial and economic environment under which the cash flows of a contract should be evaluated. Important risk factors are interest rates, foreign exchange rates, prices of shares and commodities but also credit ratings and default probabilities. Important is that their future state is unknown and has to be "guessed" in What-If calculations or randomly drawn in Monte-Carlo simulations.

This input data is used for calculating different analytic results such as liquidity or solvency of a company.

Figure 2 shows the data flow of the implemented system architecture based on Apache Spark. The two main input data sets are the risk factor scenarios and the financial contracts, from which the contract-risk factor input table is produced by building the outer product. All resulting contract-risk factor pairs need additional inputs such as reporting currency and time-period specification for liquidity aggregation. After the cash flows being calculated, the financial analytics is performed, which yields contract-level results.

## III. EXPERIMENTS

All experiments were executed on Amazon Web Services running Spark 2.3 using the Java interface of Spark and

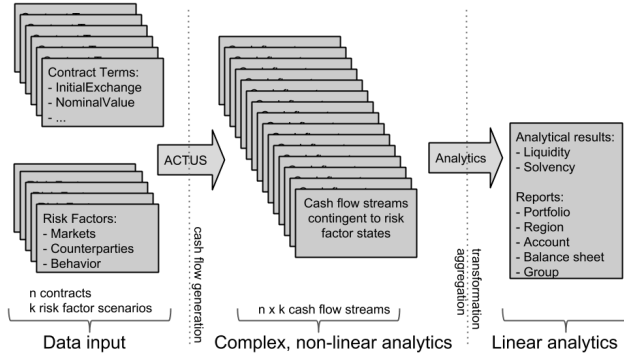


Fig. 1. Data flows in the ACTUS framework with complex, non-linear financial analytics (cash flow generation) and linear financial analytics.

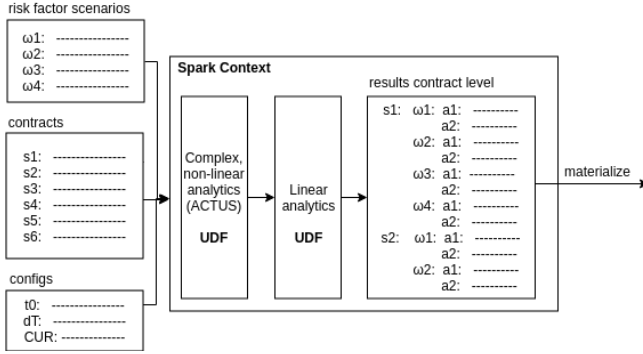


Fig. 2. System Architecture.

ACTUS. We used up to 32 instances, each with 30 GB RAM and 16 vCPU cores running at 2.5 GHz.

Our test case contains up to 96 million *financial contracts* that need 40 ACTUS *contract attributes* for input specification for the simulation and 3 additional parameters for the analytical part. The total size of the contract data set is 30 GB. In total, the largest experiments ran on 960 GB of (distributed) RAM and 512 vCPU cores.

Figure 3 shows the performance for generating cash flow events as well as calculating *nominal value*, *liquidity* and *fair value* analytics on up to 512 vCPU cores and up to 96 million contracts. For all our experiments we studied the scalability by increasing the number of CPU cores (see description of x-axis *at the bottom*) and correspondingly increasing the size of the input data set (see description of x-axis *at the top*). For instance, when doubling the number of CPU cores, we also doubled the size of the input data set. The advantage of this approach is that the performance can easily be interpreted since the ideal performance curve is a line with slope 0.

We can observe good linear scalability starting from 32 cores. We can also see that from 16 vCPU cores (which corresponds to 1 Spark worker node) to 32 vCPU cores (2 nodes), the slope of the curve is larger than 0, which indicates sub-optimal performance. The reason is that by increasing the number of nodes from 1 to 2, Spark needs to start the cluster management, which includes shuffling of data over the

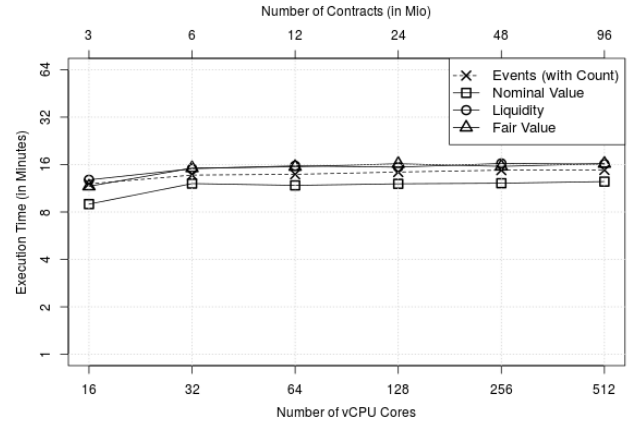


Fig. 3. Large-scale financial analytics.

network and causes an extra overhead. However, starting from 32 cores we observe almost linear scalability with a slope close to 0. Since there is no additional disproportional management overhead in our application, this is the expected behavior.

#### IV. CONCLUSIONS

In this paper we have shown that we can perform analytics of up to 96 millions financial contracts on 512 vCPU cores with almost linear scalability. Given that a large, global bank has on the order to 10 million financial contracts, these results are very promising since the system enables us to calculate core risk analytics of about 10 large banks in only about 16 minutes.

#### V. ACKNOWLEDGMENTS

The work was supported by the Swiss Commission for Technology and Innovation (CTI) under grant 25349.1 PFES-ES.

#### REFERENCES

- [1] Algorithmic Contract Type Unified Standards, Actus Financial Research Foundation, Retrieved from <https://actusfrf.org>, 2018.
- [2] Basel Committee on Banking Supervision, Principles for effective risk aggregation and risk reporting, Basel, Switzerland: Bank for International Settlement, January 2013.
- [3] W. Brummert, I. Akkizidis, W. Breyman, et al. "Unified Financial Analysis", Wiley, Chichester, 2009.
- [4] W. Breyman, N. A. Bundi, J. Heitz, J. Micheler, and K. Stockinger. Large-scale data-driven financial risk assessment, Applied Data Science, to appear.
- [5] N. Jenkinson and I. S. Leonova. "The importance of data quality for effective financial stability policiesLegal entity identifier: a first step towards necessary financial data reforms", Financial Stability Review, (17), p. 101-110, 2013.
- [6] Shareholder Report on Write-Downs, Zurich, Switzerland: UBS AG, Retrieved from <http://maths-fi.com/ubs-shareholder-report.pdf>, April 2008.